

# Dify 매뉴얼

## V1

# Dify 란?

- 오픈 소스 대규모 언어 모델(LLM) 애플리케이션 개발 플랫폼
- 개발자가 빠르게 프로덕션 수준의 생성형 AI 애플리케이션을 구축할 수 있도록 지원
  - 전문가가 아니어도 AI 애플리케이션의 정의와 데이터 운영 가능
- Dify 장점
  - 다양한 모델 지원
  - 사용하기 쉽고, 직관적인 프롬프트 오케스트레이션 인터페이스
  - 고품질 RAG 엔진
  - 프레임워크 전문가와 커뮤니티에 의해 공동 제작된 “오픈 소스”
  - 데이터 컨트롤이 용이한 유연한 보안
  - 자체 배포 가능

## !!!주의사항!!!

- 타인의 데이터나, 워크플로우를 **함부로 수정/삭제 하지 않도록 주의**
- 컴퓨팅 자원을 많이 소요하는 작업(예: 여러 개 문서 임베딩) 등은 **저녁 시간을 이용하거나, 본부 간 작업 시간 협의 후 진행**

# Contents

- 접속 및 계정 생성 방법
- 지식 생성 방법
- 워크플로우 만들기
  - Knowledge retrieval (지식 인출) 챗봇 만들기
    - 시작질문 입력, 꼬리 질문 만들기, 출처 표기
    - 배포 하기
  - 질문 분류 적용 챗봇 만들기

# 접속 및 계정 생성 방법

# 접속 및 계정 생성 방법

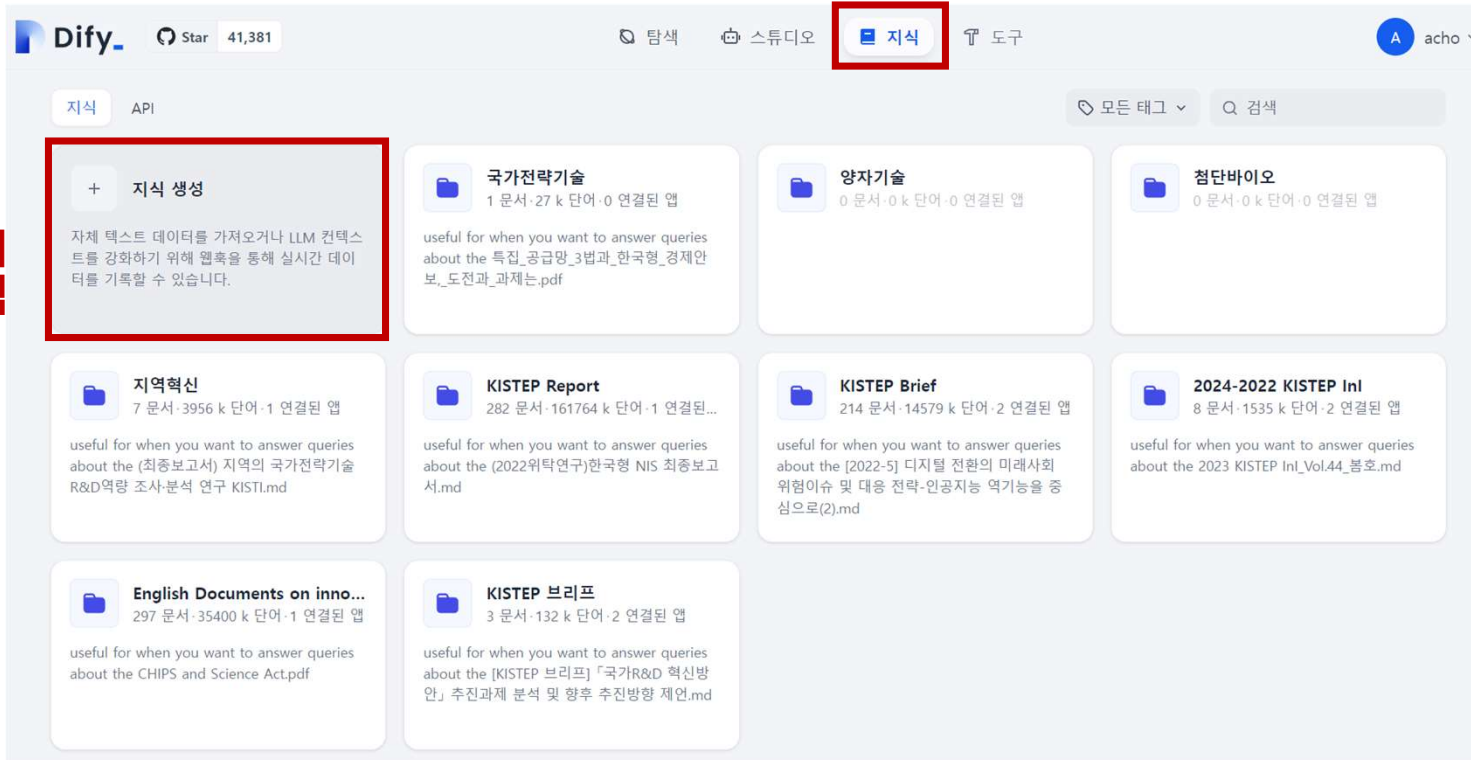
- 접속 방법 : <http://10.1.10.238:8089/>
  - 원내 접속 (외부 접속 시 VPN 사용 필요)
- 계정 생성을 위해서는 초대 링크 발송 필요
  - 전략기획센터 조아라([acho@kistep.re.kr](mailto:acho@kistep.re.kr), 2511)에게 요청
  - 문서 업로드 및 AI 서비스 개발(스튜디오 사용)을 위해서는 Editor(편집자) 권한 필요

## 지식 생성 방법

- 답변 생성 시 참조할 지식 꾸러미를 생성
- (생성 과정) 문서들 업로드 → 적당한 크기로 chunking → 임베딩

# 지식 생성 방법

## 1. [지식] 클릭



## 2. [지식 생성] 클릭



# 지식 생성 방법

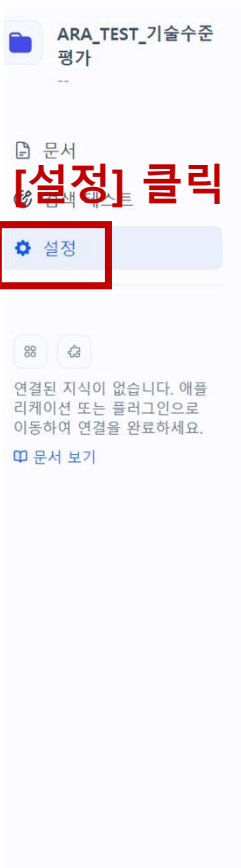
- 지식 생성 시 문서를 업로드 하거나
- 비어 있는 지식 생성 후 추후 문서 업로드

The screenshot displays the '지식 생성' (Knowledge Generation) interface. On the left, a sidebar lists the steps: 1. 데이터 소스 선택 (Data Source Selection), 2. 텍스트 전처리 및 클리닝 (Text Preprocessing and Cleaning), and 3. 실행 및 완료 (Execution and Completion). The main area is titled '데이터 소스 선택' (Data Source Selection) and offers three options: '텍스트 파일에서 가져오기' (Import from Text File), 'Notion 동기화' (Sync with Notion), and '웹 사이트 동기화' (Sync with Website). The 'Import from Text File' option is highlighted with a red box. Below it, the '텍스트 파일 업로드' (Text File Upload) section is also highlighted with a red box, containing an upload icon and the text '파일을 끌어다 놓거나 찾아보기' (Drag and drop files or browse). Below this, a blue button labeled '다음' (Next) is followed by the text '지식 생성 시 문서 업로드' (Upload document during knowledge generation). Below that, a button labeled '비어있는 지식 생성' (Create empty knowledge) is highlighted with a red box. A red arrow points from this button to a modal window titled '비어있는 지식 생성' (Create empty knowledge). The modal contains a close button (X), a message stating that no documents are included in empty knowledge and they can be uploaded later, a text input field for '지식 이름' (Knowledge Name) with the value 'ARA\_TEST', and '취소' (Cancel) and '생성' (Generate) buttons.

또는 비어 있는 지식 생성 후 추후 업로드

# 지식 생성 방법

## 1. [설정] 클릭



### 지식 설정

여기에서 지식의 속성과 작동 방법을 변경할 수 있습니다.

## 2. 지식 이름 및 설명 입력

지식 이름: ARA\_TEST\_기술수준평가

지식 설명: 이 지식에 포함된 내용을 설명하세요. 자세한 설명은 AI가 지식 내용에 빠르게 접근할 수 있도록 합니다. 비어 있으면 Dify가 기본 검색 전략을 사용합니다.

좋은 지식 설명 작성 방법 배우기

### 권한

권한: 나만

## 3. 권한 설정

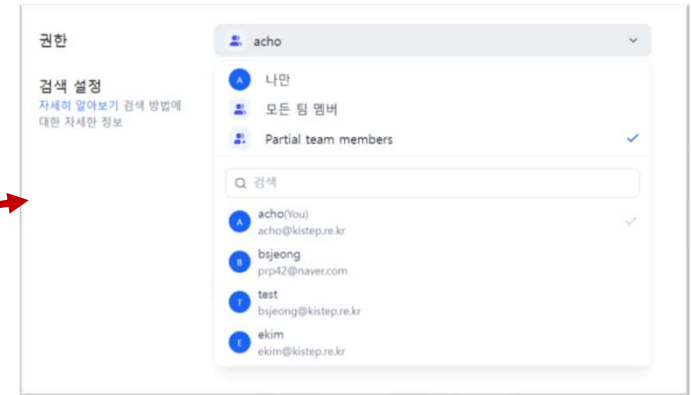
### 검색 설정

자세히 알아보기 검색 방법에 대한 자세한 정보

역 인덱스: 효율적인 검색에 사용되는 구조입니다. 각 용어는 문서나 웹 페이지에 포함된 것을 가리키며, 용어마다 체계적으로 정리되어 있습니다.

상위 K: 2

저장



- 나만
- 모든 팀 멤버
  - 모든 사람이 이 문서를 열람/수정하길 원할 경우 선택
- Partial team member (허가받은 사람만)
  - 특정인 (특정 부서 사람들)이 이 문서를 열람/수정하길 원할 경우 선택

# 지식 생성 방법

- 현재 문서는 **15M까지만 업로드 가능**
  - (50MB로 안내되어 있으나 15M 이상 업로드 시 여전히 오류, 추후 업데이트 요청 예정)

← 데이터 추가

1 데이터 소스 선택

2 텍스트 전처리 및 클리닝

3 실행 및 완료

### 데이터 소스 선택

텍스트 파일에서 가져오기 Notion 동기화 웹 사이트 동기화

### 텍스트 파일 업로드

파일을 끌어다 놓거나 찾아보기

TXT, MARKDOWN, PDF, HTML, XLSX, XLS, DOCX, CSV을(를) 지원합니다. 파일당 최대 크기는 15MB입니다.

★(최종)과학기술예측조사+실현도+평가를+통한+예측방법론+고도화+연구-4교.pdf 5.13MB

**다음** [다음] 클릭

# 지식 생성 방법

- 다수의 자료를 임베딩 할 때에는
  - 저녁 시간을 이용하거나
  - 본부 별 사용 시간 협의 후 진행

The interface shows several configuration panels:

- 텍스트 전처리 및 클리닝**: Includes '체크 설정' (자동) and '사용자 설정'.
- 인덱스 모드**: A red box highlights the '고품질 추천' (High Quality Recommendation) mode. Text next to it says '고품질로 임베딩 해도 과금되지 않음' (Embedding with high quality does not incur charges).
- 질문과 답변 형식으로 세그먼트화**: A toggle switch for segmenting by question and answer format.
- 검색 설정**: Includes '벡터 검색' (Vector Search) and '문서 전처리' (Document Preprocessing).
- 문서 전처리**: A red box highlights the '클릭' (Click) button and the '저장하고 처리' (Save and Process) button.

The document list shows the following details:

#	파일명	단어 수	검색 횟수	업로드 시간	상태	동작
1	★(최종)과학기술예측조사+실현...	97.3k	0	2024/08/13 10:29	색인화 중	...

The document list shows the following details:

#	파일명	단어 수	검색 횟수	업로드 시간	상태	동작
1	★(최종)과학기술예측조사+실현...	97.3k	11	2024/08/13 10:29	사용 가능	...

완료

## 워크플로우 만들기

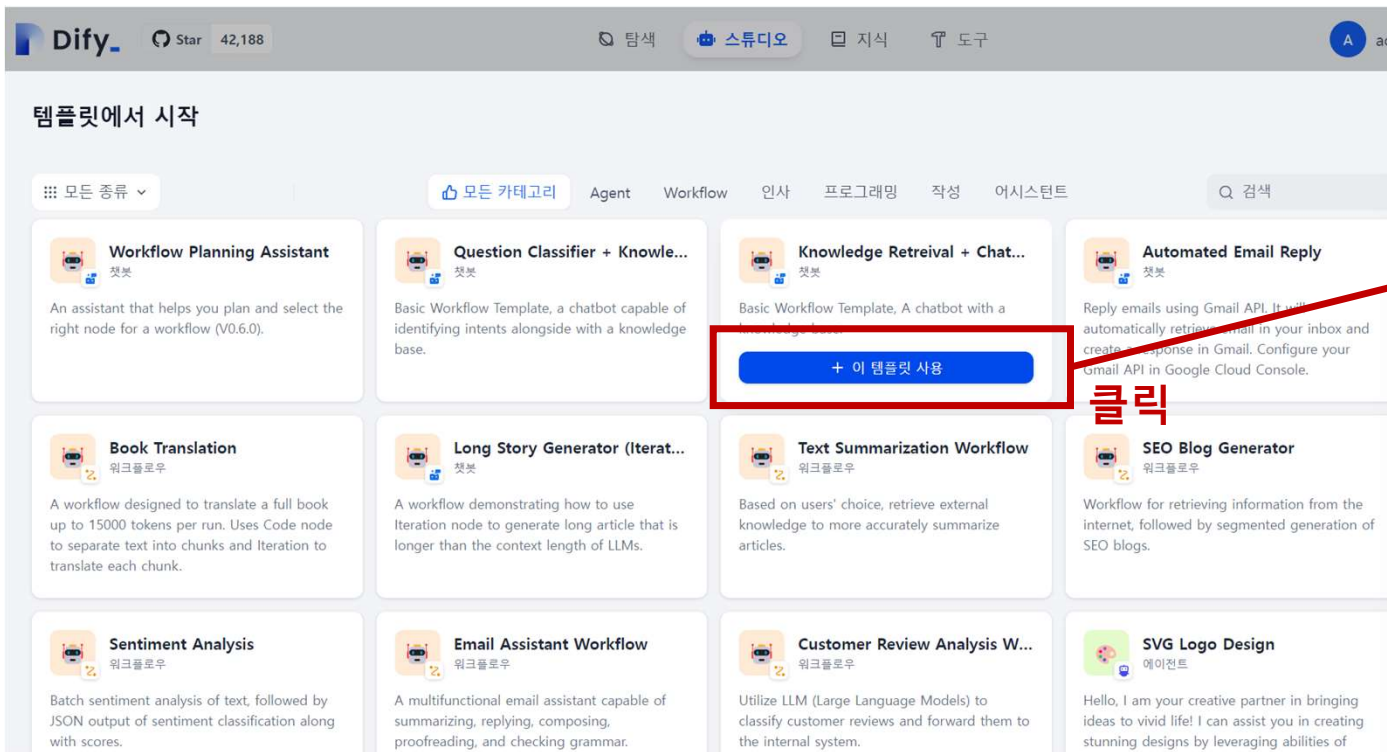
- Knowledge retrieval (지식 인출) 챗봇 만들기
- 질문 분류 적용 챗봇 만들기

# Knowledge retrieval (지식 인출) 챗봇 만들기

- 생성한 지식을 참조하여 답변을 생성하는 챗봇을 만들고자 함
- (과정) 1.질문 입력 → 2.생성한 지식에서 관련 정보(chunk)를 인출 → 3.인출한 정보(chunk)를 참조하여 답변 생성
- 템플릿에서 시작하면 편리함

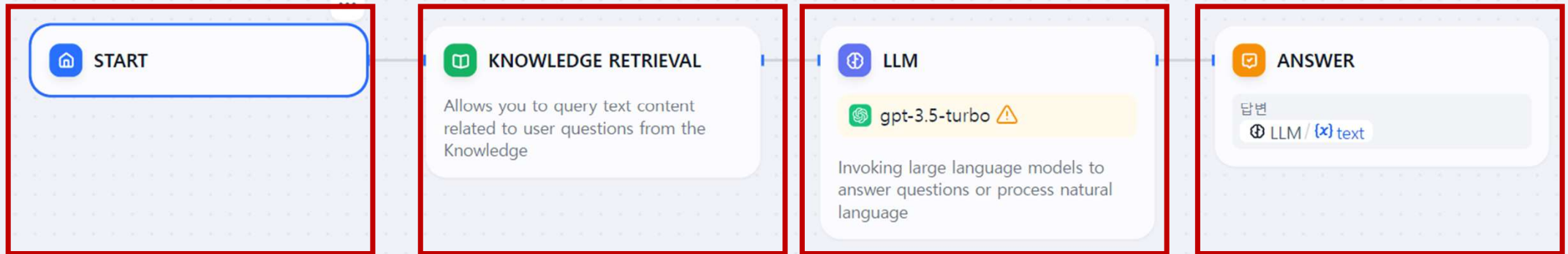


# Knowledge retrieval (지식 인출) 챗봇 만들기



# Knowledge retrieval (지식 인출) 챗봇 만들기

- Knowledge retrieval + chatbot 의 기본값은 아래와 같음
- 아래와 같은 과정으로 답변이 생성됨
  - 본 예제에서는 2번과 3번 과정을 적절히 커스터마이징하여 챗봇을 생성할 예정



1. 시작(질문)

2. 질문에 적합한 지식  
인출

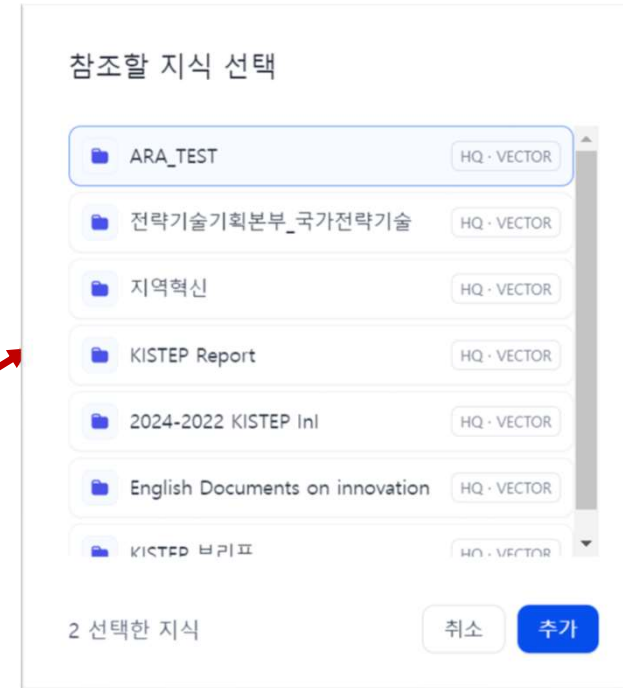
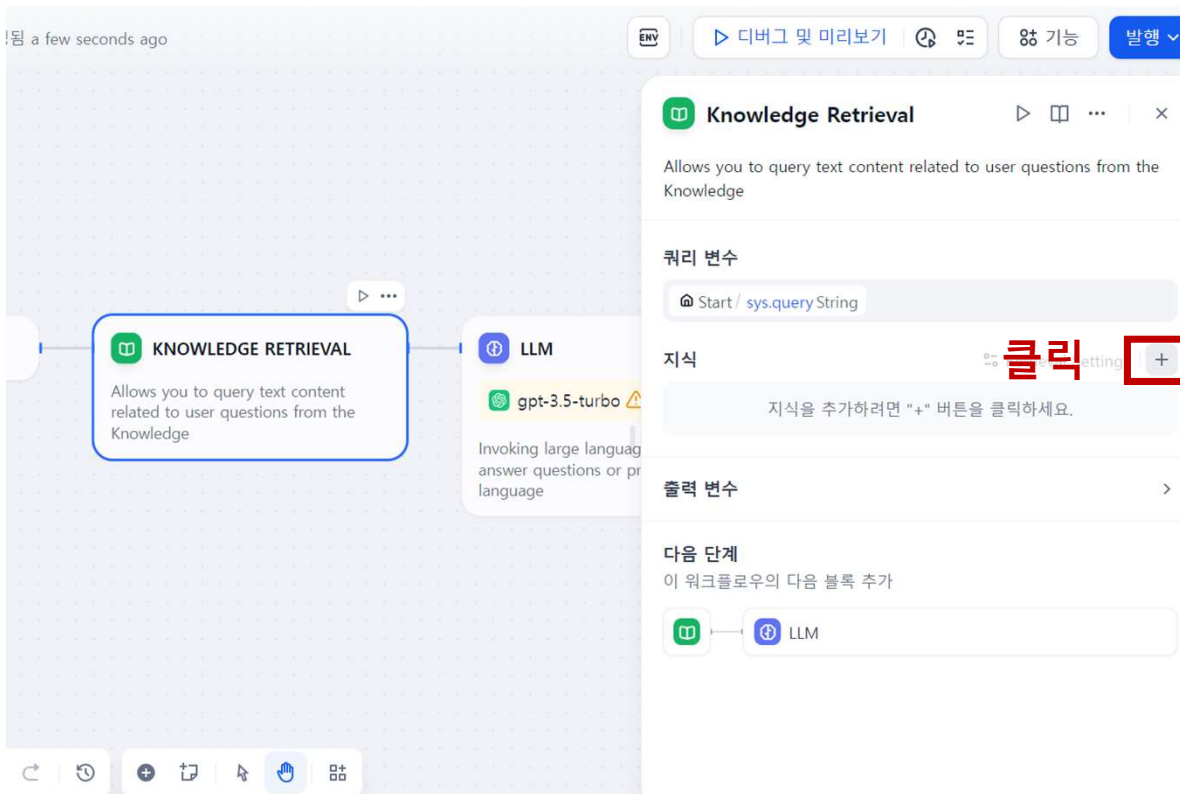
3. 인출된 지식을 참조  
하여 답변 생성

4. 답변



# Knowledge retrieval (지식 인출) 챗봇 만들기

- Knowledge retrieval 커스터마이징 하는 법
  1. [+] 버튼 클릭하여 지식 추가
  2. [Retrieval Setting] 버튼 클릭하여 언어모델(시스템 추론 모델) 선택

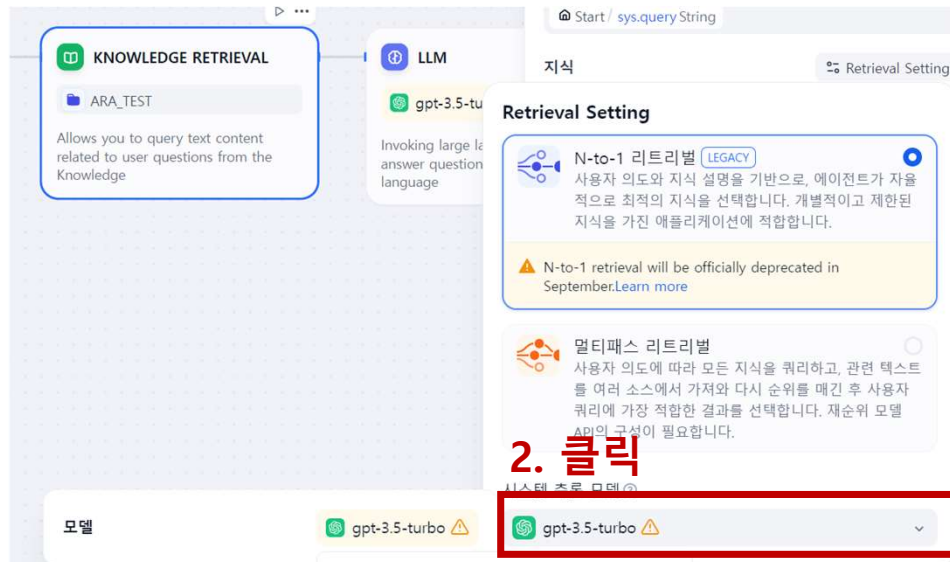


답변 생성 시  
참조할 지식 선택  
(복수 선택 가능)

# Knowledge retrieval (지식 인출) 챗봇 만들기

## • Knowledge retrieval 커스터마이징 하는 법

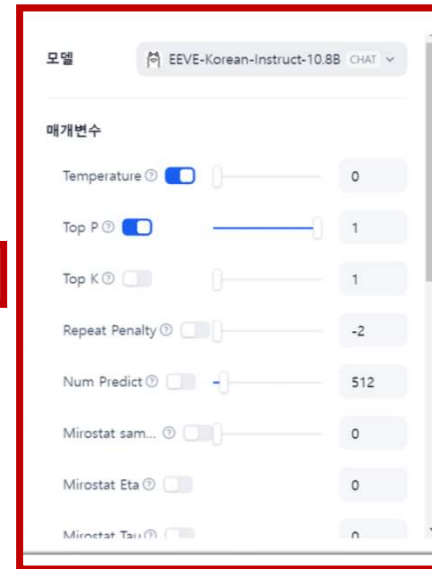
1. [+] 버튼 클릭하여 지식 추가
2. [Retrieval Setting] 버튼 클릭하여 언어모델(시스템 추론 모델) 선택



2. 클릭



1. 클릭

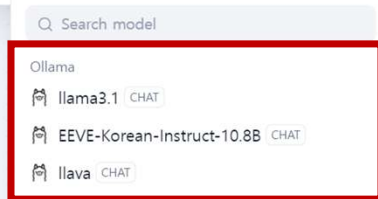


## 4. 파라미터 조정

- Temperature : 모델이 다음 토큰 후보 중에서 출력할 토큰을 선택하는 방식
  - 높을수록 : 창의적인 답변(무작위적, 환각 많음)
  - 낮을수록 : 일관된 답변(출력을 높은 확률의 토큰에 집중)
- Top P, Top K : 확률이 P%인 토큰, 혹은 확률이 상위 K개인 토큰의 결과를 출력 후보로 지정
  - 높을수록 diverse text
  - 낮을수록 focused and conservative text

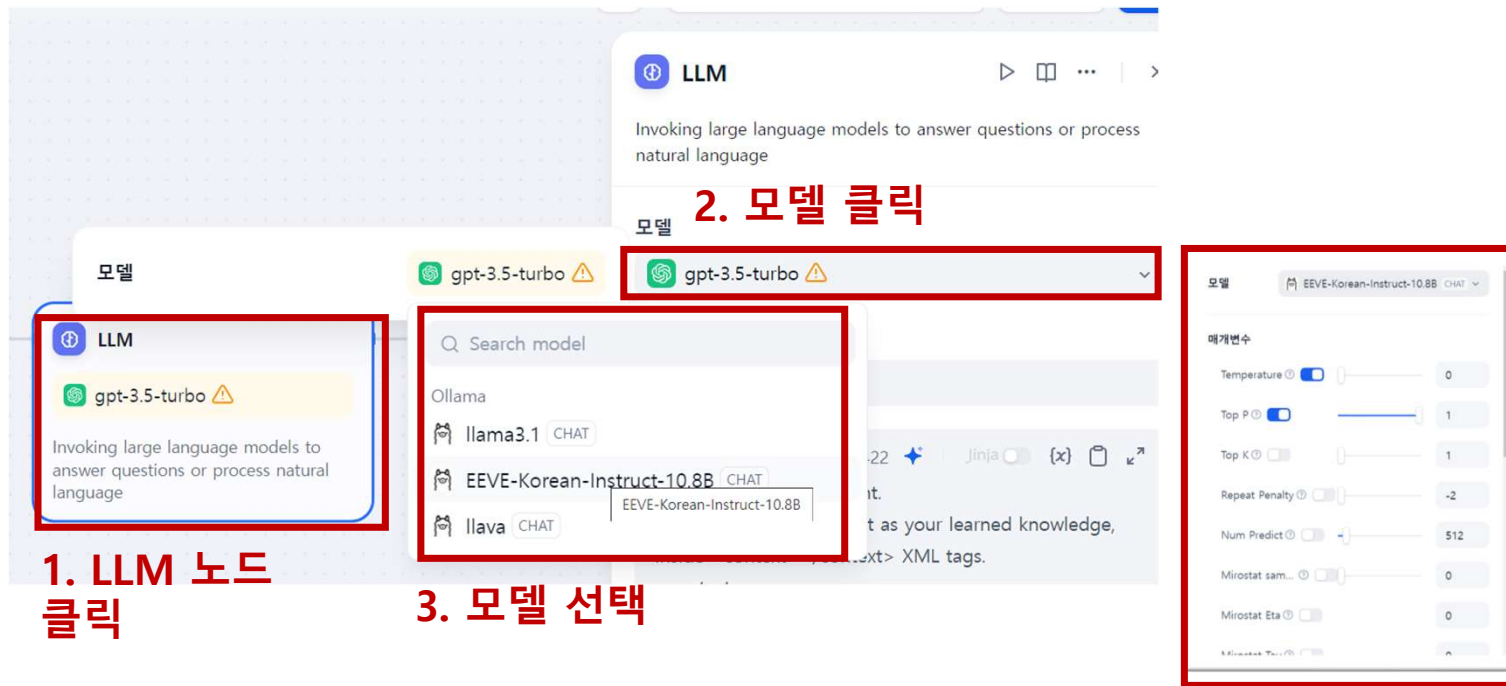
## 3. 모델 선택

- 현재 GPT 모델은 사용 불가
- 한글의 경우 : EEVE 선택
- 영어의 경우 : llama3.1 선택



# Knowledge retrieval (지식 인출) 챗봇 만들기

- LLM 모델 선택



## 4. 파라미터 조정

# Knowledge retrieval (지식 인출) 챗봇 만들기

## • 컨텍스트 및 시스템 프롬프트 편집

### 컨텍스트 ⓘ 1. 컨텍스트 클릭

Knowledge Retrieval / result Array[Object]

변수 검색

START

- (x) sys.query String
- (x) sys.conversation\_id String
- (x) sys.user\_id String

KNOWLEDGE RETRIEVAL

- (x) result Array[Object]

### 2. 앞에서 지정한 지식을 컨텍스트로 선택

SYSTEM ⓘ 422 Jinja ⓘ (x) ⓘ

You are a helpful assistant.

Use the following context as your learned knowledge, inside <context> </context> XML tags.

<context>

컨텍스트

</context>

When answer to user:

- If you don't know, just say that you don't know.
- If you don't know when you are not sure, ask for clarification.

Avoid mentioning that you obtained the information from the context.

And answer according to the language of the user's question.

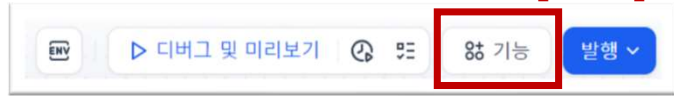
+ 메시지 추가

### 3. 필요시 system prompt 편집

# Knowledge retrieval (지식 인출) 챗봇 만들기

- 시작질문 입력, 꼬리 질문 만들기, 출처 표기

1. 우측 상단 [기능] 클릭



2. 좌측 상단 [기능추가] 클릭



3. 시작질문 입력 활성화



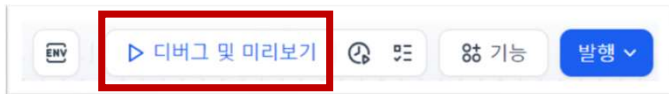
4. 꼬리 질문 만들기

5. 출처 표기

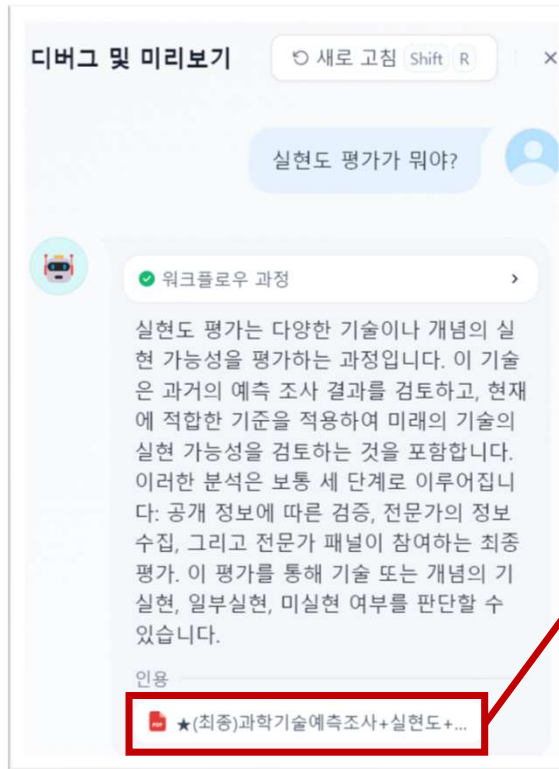


# Knowledge retrieval (지식 인출) 챗봇 만들기

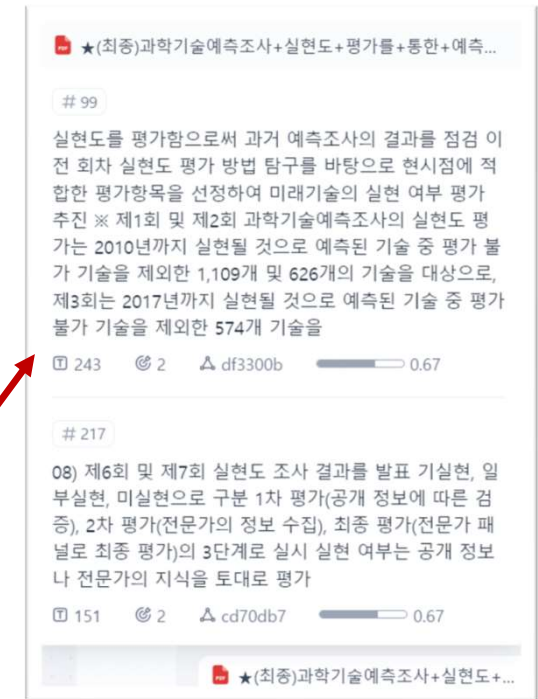
- 실행해 보기 (디버그 및 미리보기)



1. 우측 상단 [디버그 및 미리보기] 클릭



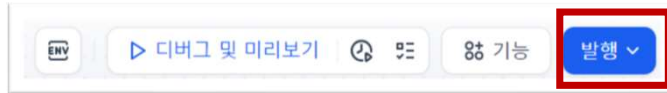
2. 질문 입력 및 답변 확인



3. 답변 생성된 지식(chunk 확인)

# Knowledge retrieval (지식 인출) 챗봇 만들기

## • 배포



1. 우측 상단 [배포] 클릭

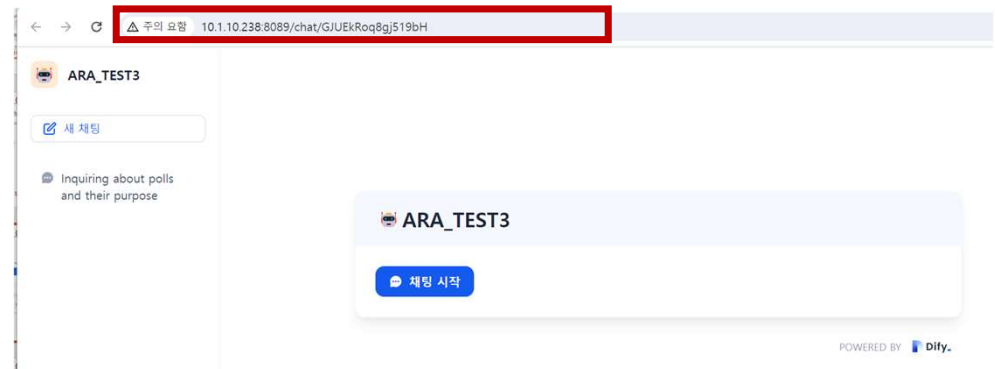


2. [업데이트] 클릭



3. [앱 실행] 클릭

4. 새 창에 뜨는 URL 배포



## 워크플로우 만들기

- Knowledge retrieval (지식 인출) 챗봇 만들기
- 질문 분류 적용 챗봇 만들기

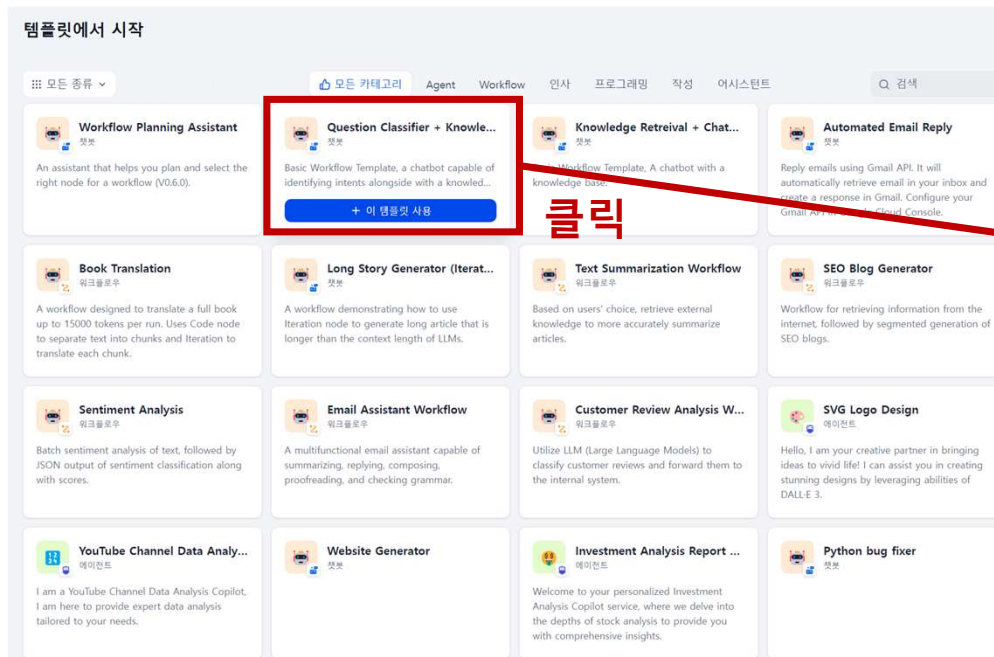


## !!!주의!!!

- 디폴트 값을 사용하고, 클래스를 대강 입력하였을 경우, **분기가 적절히 진행되지 않을 수 있음**
- 분류를 잘 하기 위해서는 **파라미터 조정, 클래스 설명 최적화 등 시행착오가 필요함**
- **지식을 인출하지 않고 LLM에서 답변을 하는 경우가 종종 있으며, 이 또한 파라미터, 프롬프트 조정 등을 위한 최적화가 필요할 것으로 보임**

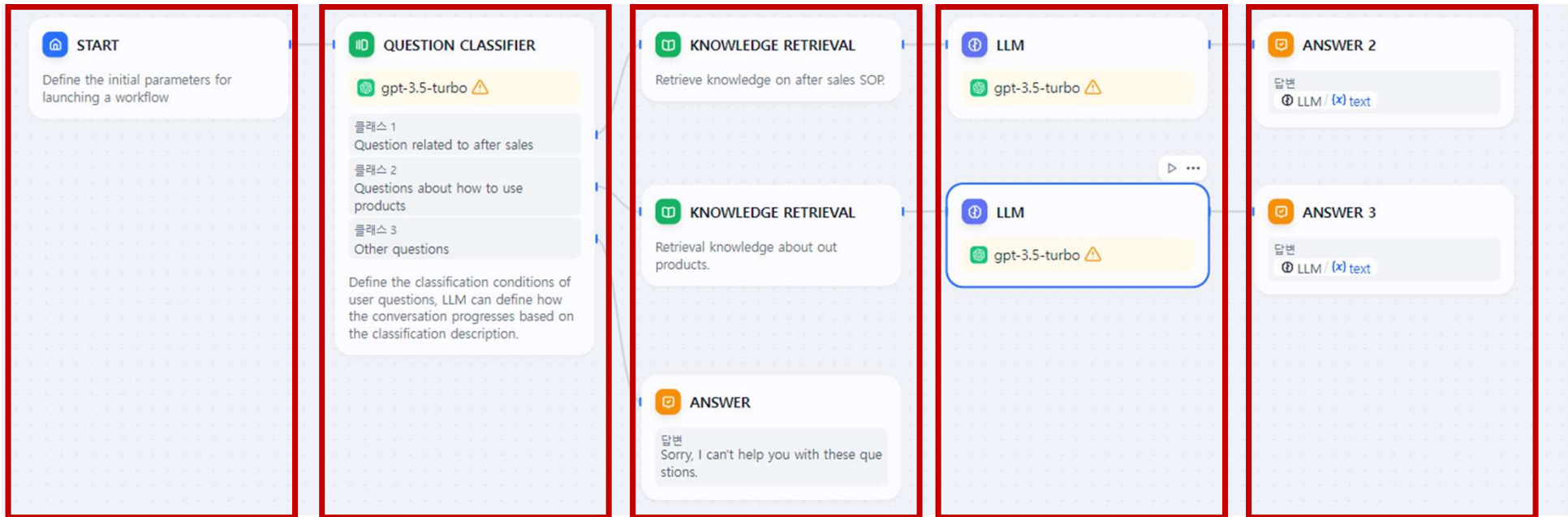
# 질문 분류 적용 챗봇 만들기

- Knowledge retrieval(지식 인출) 챗봇 만들기에 사용자가 입력한 질문을 분류하는 단계를 추가
- (과정) 1.질문 입력 → 2.사용자 질문 분류 → 3. 분류된 질문에 대응하는 지식에서 관련 정보(chunk)를 인출 → 3.인출한 정보(chunk)를 참조하여 답변 생성
- 템플릿에서 시작하면 편리함



# 질문 분류 적용 챗봇 만들기

- Knowledge retrieval + chatbot 의 기본값은 아래와 같음
- 아래와 같은 과정으로 답변이 생성됨



1. 시작(질문)

2. 질문을 서로 다른 클래스로 분류

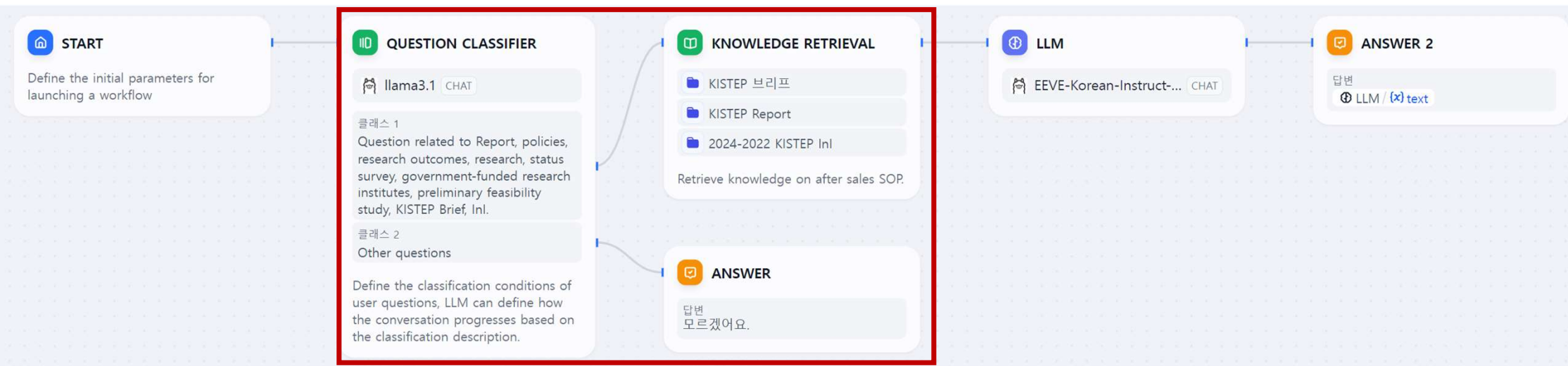
3. '분류한 클래스에 맞는 지식을 인출' 또는 '인출 없이 답변 후 종료'

4. 인출된 지식을 참조하여 답변 생성

5. 답변

# 질문 분류 적용 챗봇 만들기

- 본 예제에서는 2개의 클래스만 생성해 봄
  - 클래스는 추가 가능



질문을 두 클래스로 분류

1. 과기정책 관련 질문일 경우 지식을 참조하여 답변을 생성
2. 다른 질문일 경우 “모르겠어요” 라고 답변

# 질문 분류 적용 챗봇 만들기

## • 클래스 입력

### 1. Question classifier 노드 선택

**QUESTION CLASSIFIER**

llama3.1 CHAT

클래스 1  
Question related to Report, policies, research outcomes, research, status survey, government-funded research institutes, preliminary feasibility study, KISTEP Brief, Inl.

클래스 2  
Other questions

Define the classification conditions of user questions, LLM can define how the conversation progresses based on the classification description.

### 2. 언어모델은 llama3.1 선택 (EVEE 모델은 분류작업을 잘 못함)

llama3.1 CHAT

클래스

클래스	개수	삭제	복사	확대
클래스 1 Question related to Report, policies, research outcomes, research, status survey, government-funded research	170			
클래스 2 Other questions	15			

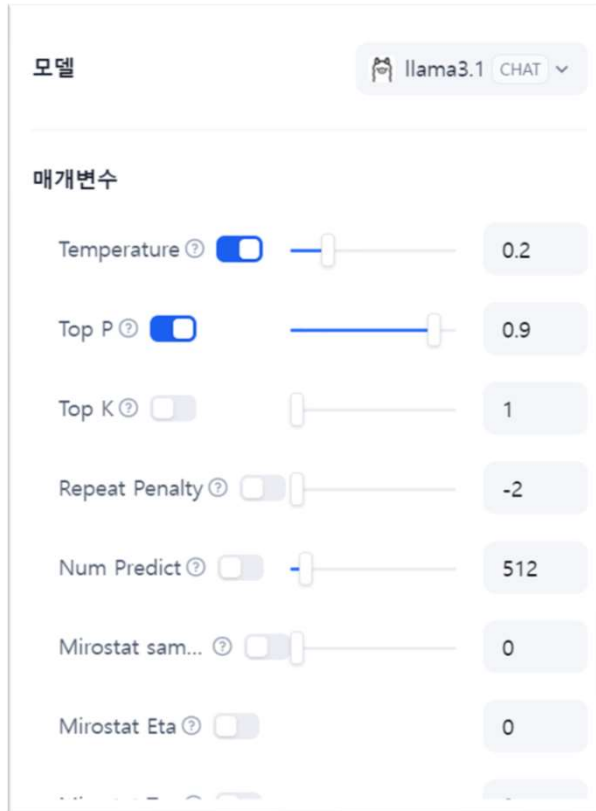
+ 클래스 추가

### 3. 클래스 설명 입력

고급 설정 >

출력 변수 >

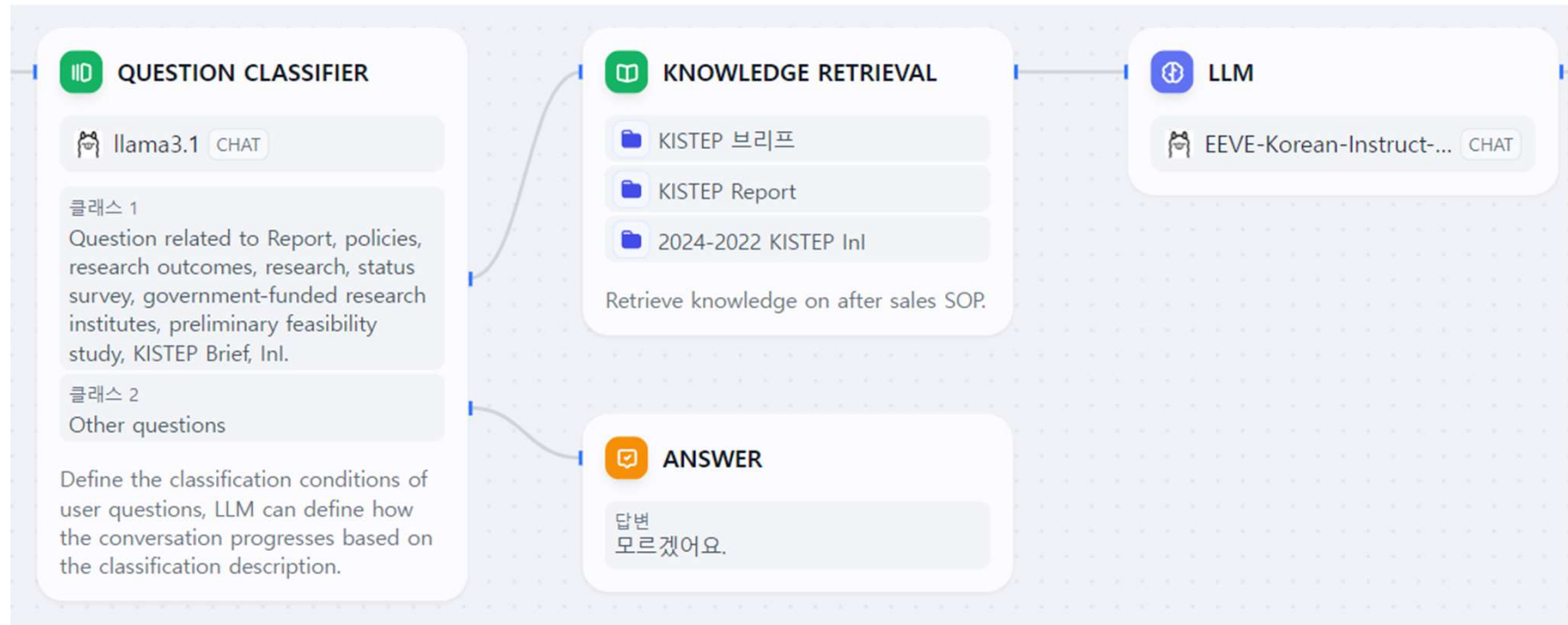
# 질문 분류 적용 챗봇 만들기



**!!!주의!!!**

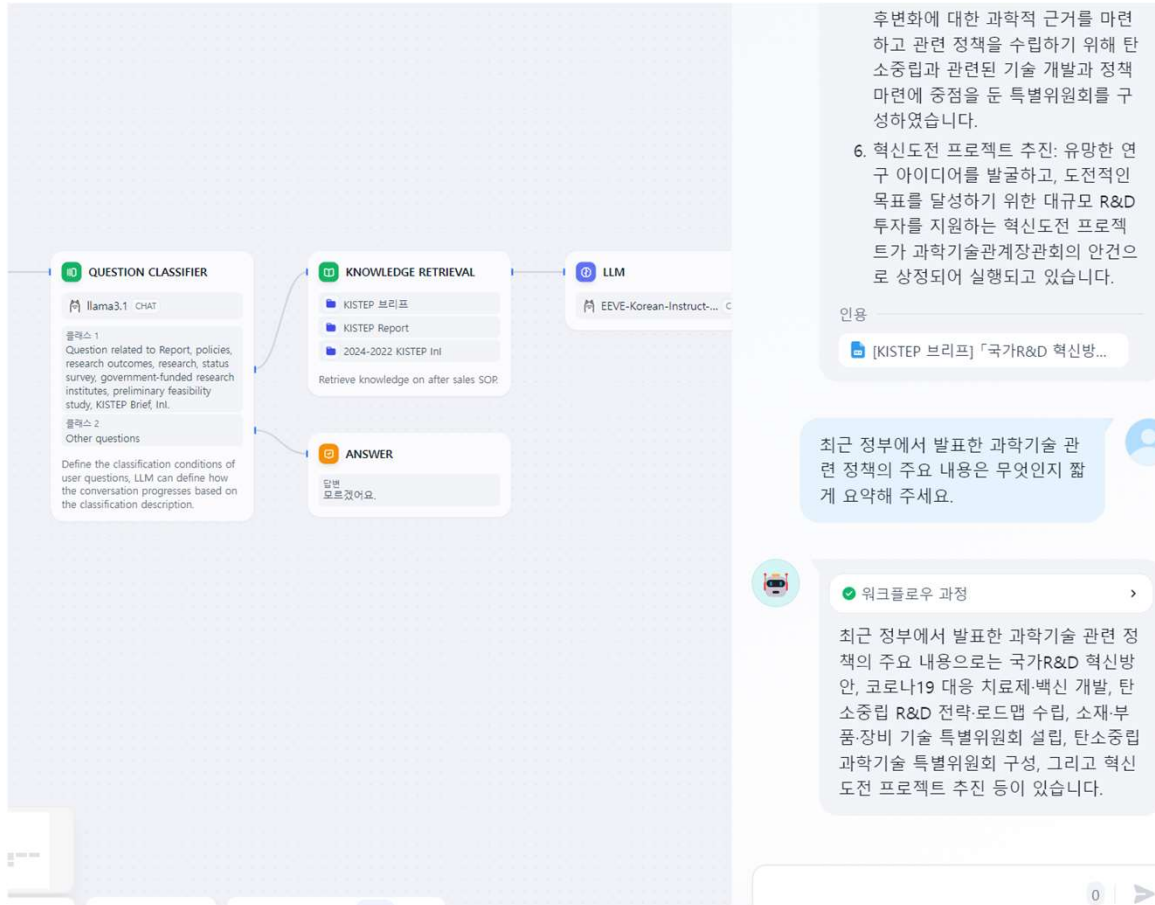
- **Classification의 경우 매개변수를 잘 조정해야 함.**
- **변경 후 테스트 작업을 반복해 보며 조절 필요.**

# 질문 분류 적용 챗봇 만들기



나머지 Knowledge Retrieval, LLM은 앞 부분과 같음  
단, LLM 파라미터(매개변수)도 시행착오를 거쳐 조정할 필요 있음

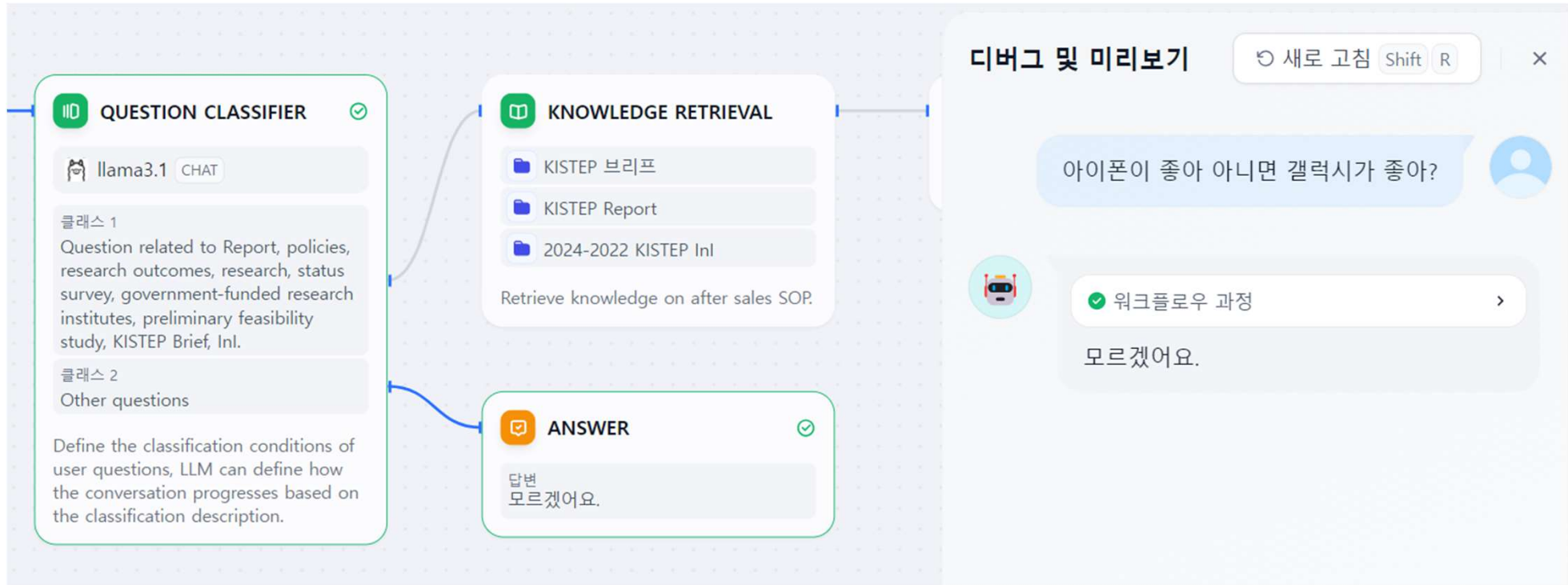
# 질문 분류 적용 챗봇 만들기



- 과기정책과 관련된 질문을 하면 답변 생성함
- [디버그 및 미리보기] 실행 시 왼쪽에 활성화 되는 노드 확인 가능



# 질문 분류 적용 챗봇 만들기



- **관련 없는 질문 할 경우 모르겠다고 답변**